

Final Project

2025-07-12

```
library(readr)
proj_genes <- read_csv("/Users/ayer/Desktop/QBS103_GSE157103_genes.csv")
```

```
## New names:
## Rows: 100 Columns: 127
## -- Column specification
## ----- Delimiter: "," chr
## (1): ...1 dbl (126): COVID_01_39y_male_NonICU, COVID_02_63y_male_NonICU,
## COVID_03_33y_...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
library(readr)
proj_matrix <- read_csv("/Users/ayer/Desktop/QBS103_GSE157103_series_matrix-1.csv")
```

```
## Rows: 126 Columns: 25
## -- Column specification -----
## Delimiter: ","
## chr (21): participant_id, geo_accession, status, !Sample_submission_date, la...
## dbl (4): channel_count, charlson_score, ventilator-free_days, hospital-free...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Because I need to match the participant_id column/row position in order to link the two data sets, I

Converting raw data to data frame

```
proj_genes_df <- as.data.frame(proj_genes)
```

Setting the row names as the gene names

```
rownames(proj_genes_df) <- proj_genes_df[[1]]
```

Now I have to remove the first column since I am using them as row names

```
proj_genes_df <- proj_genes_df[, -1]
```

Transposing the data so that the gene names go across the top and the participant ID's became the col

```
new_proj_genes <- as.data.frame(t(proj_genes_df))
```

Just double checking

```
colnames(proj_matrix)
```

```
## [1] "participant_id"
## [2] "geo_accession"
## [3] "status"
## [4] "!Sample_submission_date"
## [5] "last_update_date"
```

```
## [6] "type"
## [7] "channel_count"
## [8] "source_name_ch1"
## [9] "organism_ch1"
## [10] "disease_status"
## [11] "age"
## [12] "sex"
## [13] "icu_status"
## [14] "apacheii"
## [15] "charlson_score"
## [16] "mechanical_ventilation"
## [17] "ventilator-free_days"
## [18] "hospital-free_days_post_45_day_followup"
## [19] "ferritin(ng/ml)"
## [20] "crp(mg/l)"
## [21] "ddimer(mg/l_feu)"
## [22] "procalcitonin(ng/ml):"
## [23] "lactate(mmol/l)"
## [24] "fibrinogen"
## [25] "sofa"
```

```
colnames(new_proj_genes)
```

```
## [1] "A1BG"      "A1CF"      "A2M"      "A2ML1"     "A3GALT2"
## [6] "A4GALT"    "A4GNT"    "AAAS"     "AACS"      "AADAC"
## [11] "AADACL2"   "AADACL3"   "AADACL4"   "AADAT"     "AAGAB"
## [16] "AAK1"      "AAMDC"     "AAMP"      "AANAT"     "AAR2"
## [21] "AARD"      "AARS1"     "AARS2"     "AARSD1"    "AASDH"
## [26] "AASDHPPT"  "AASS"      "AATF"      "AATK"      "ABAT"
## [31] "ABCA1"     "ABCA10"    "ABCA12"    "ABCA13"    "ABCA2"
## [36] "ABCA3"     "ABCA4"     "ABCA5"     "ABCA6"     "ABCA7"
## [41] "ABCA8"     "ABCA9"     "ABCB1"     "ABCB10"    "ABCB11"
## [46] "ABCB4"     "ABCB5"     "ABCB6"     "ABCB7"     "ABCB8"
## [51] "ABCB9"     "ABCC1"     "ABCC10"    "ABCC11"    "ABCC12"
## [56] "ABCC2"     "ABCC3"     "ABCC4"     "ABCC5"     "ABCC6"
## [61] "ABCC8"     "ABCC9"     "ABCD1"     "ABCD2"     "ABCD3"
## [66] "ABCD4"     "ABCE1"     "ABCF1"     "ABCF2"     "ABCF2-H2BE1"
## [71] "ABCF3"     "ABCG1"     "ABCG2"     "ABCG4"     "ABCG5"
## [76] "ABCG8"     "ABHD1"     "ABHD10"    "ABHD11"    "ABHD12"
## [81] "ABHD12B"   "ABHD13"    "ABHD14A"   "ABHD14A-ACY1" "ABHD14B"
## [86] "ABHD15"    "ABHD16A"   "ABHD16B"   "ABHD17A"   "ABHD17B"
## [91] "ABHD17C"   "ABHD18"    "ABHD2"     "ABHD3"     "ABHD4"
## [96] "ABHD5"     "ABHD6"     "ABHD8"     "ABI1"      "ABI2"
```

```
# Now I am adding a new column for "participant_id" to "new_proj_genes" with the row names, so that when
new_proj_genes$participant_id <- rownames(new_proj_genes)
```

```
# Merging the two datasets
```

```
merged_proj_data <- merge(proj_matrix, new_proj_genes, by = "participant_id")
```

```
# Double checking to see if the merge was successful
```

```
summary(merged_proj_data)
```

```
## participant_id      geo_accession      status
## Length:126          Length:126          Length:126
## Class :character    Class :character    Class :character
```

```

## Mode :character Mode :character Mode :character
##
##
##
## !Sample_submission_date last_update_date type channel_count
## Length:126 Length:126 Length:126 Min. :1
## Class :character Class :character Class :character 1st Qu.:1
## Mode :character Mode :character Mode :character Median :1
## Mean :1
## 3rd Qu.:1
## Max. :1
## source_name_ch1 organism_ch1 disease_status age
## Length:126 Length:126 Length:126 Length:126
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## sex icu_status apacheii charlson_score
## Length:126 Length:126 Length:126 Min. : 0.0
## Class :character Class :character Class :character 1st Qu.: 2.0
## Mode :character Mode :character Mode :character Median : 3.0
## Mean : 3.5
## 3rd Qu.: 5.0
## Max. :11.0
## mechanical_ventilation ventilator-free_days
## Length:126 Min. : 0.00
## Class :character 1st Qu.:12.75
## Mode :character Median :28.00
## Mean :20.35
## 3rd Qu.:28.00
## Max. :28.00
## hospital-free_days_post_45_day_followup ferritin(ng/ml) crp(mg/l)
## Min. : 0.00 Length:126 Length:126
## 1st Qu.: 0.00 Class :character Class :character
## Median :29.50 Mode :character Mode :character
## Mean :24.19
## 3rd Qu.:39.00
## Max. :44.00
## ddimer(mg/l_feu) procalcitonin(ng/ml): lactate(mmol/l) fibrinogen
## Length:126 Length:126 Length:126 Length:126
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## sofa A1BG A1CF A2M
## Length:126 Min. :0.0500 Min. :0.000000 Min. :0.0000
## Class :character 1st Qu.:0.2100 1st Qu.:0.000000 1st Qu.:0.0400
## Mode :character Median :0.3100 Median :0.000000 Median :0.0900
## Mean :0.3624 Mean :0.002222 Mean :0.1237
## 3rd Qu.:0.4200 3rd Qu.:0.000000 3rd Qu.:0.1500
## Max. :2.8000 Max. :0.020000 Max. :0.8800
## A2ML1 A3GALT2 A4GALT A4GNT

```

| | | | | |
|----|------------------|-------------------|-----------------|------------------|
| ## | Min. :0.00000 | Min. :0.0000 | Min. :0.00000 | Min. :0.00000 |
| ## | 1st Qu.:0.00000 | 1st Qu.:0.0000 | 1st Qu.:0.00000 | 1st Qu.:0.00000 |
| ## | Median :0.01000 | Median :0.0500 | Median :0.00000 | Median :0.00000 |
| ## | Mean :0.01413 | Mean :0.1292 | Mean :0.01556 | Mean :0.01254 |
| ## | 3rd Qu.:0.02000 | 3rd Qu.:0.1675 | 3rd Qu.:0.00000 | 3rd Qu.:0.00000 |
| ## | Max. :0.10000 | Max. :1.7500 | Max. :0.72000 | Max. :0.19000 |
| ## | AAAS | AACS | AADAC | AADACL2 |
| ## | Min. : 5.17 | Min. :0.350 | Min. :0 | Min. :0.000000 |
| ## | 1st Qu.:11.48 | 1st Qu.:1.613 | 1st Qu.:0 | 1st Qu.:0.000000 |
| ## | Median :15.57 | Median :2.210 | Median :0 | Median :0.000000 |
| ## | Mean :16.24 | Mean :2.356 | Mean :0 | Mean :0.001667 |
| ## | 3rd Qu.:19.79 | 3rd Qu.:2.995 | 3rd Qu.:0 | 3rd Qu.:0.000000 |
| ## | Max. :29.46 | Max. :5.630 | Max. :0 | Max. :0.030000 |
| ## | AADACL3 | AADACL4 | AADAT | AAGAB |
| ## | Min. :0.000000 | Min. :0.0000000 | Min. :0.00000 | Min. : 4.55 |
| ## | 1st Qu.:0.000000 | 1st Qu.:0.0000000 | 1st Qu.:0.00000 | 1st Qu.:11.69 |
| ## | Median :0.000000 | Median :0.0000000 | Median :0.00000 | Median :14.95 |
| ## | Mean :0.002063 | Mean :0.0005556 | Mean :0.00619 | Mean :15.83 |
| ## | 3rd Qu.:0.000000 | 3rd Qu.:0.0000000 | 3rd Qu.:0.00000 | 3rd Qu.:18.42 |
| ## | Max. :0.070000 | Max. :0.0400000 | Max. :0.06000 | Max. :30.85 |
| ## | AAK1 | AAMDC | AAMP | AANAT |
| ## | Min. : 0.840 | Min. : 4.890 | Min. : 8.59 | Min. :0.0000 |
| ## | 1st Qu.: 3.790 | 1st Qu.: 9.072 | 1st Qu.:24.36 | 1st Qu.:0.1000 |
| ## | Median : 5.095 | Median :12.260 | Median :33.95 | Median :0.1750 |
| ## | Mean : 5.413 | Mean :12.897 | Mean :35.60 | Mean :0.3098 |
| ## | 3rd Qu.: 6.490 | 3rd Qu.:15.510 | 3rd Qu.:46.35 | 3rd Qu.:0.3850 |
| ## | Max. :13.890 | Max. :24.820 | Max. :78.72 | Max. :1.4600 |
| ## | AAR2 | AARD | AARS1 | AARS2 |
| ## | Min. : 2.740 | Min. :0.00000 | Min. : 1.550 | Min. :0.300 |
| ## | 1st Qu.: 7.862 | 1st Qu.:0.03000 | 1st Qu.: 5.112 | 1st Qu.:0.950 |
| ## | Median :10.475 | Median :0.06500 | Median : 8.965 | Median :1.495 |
| ## | Mean :11.689 | Mean :0.08413 | Mean :10.257 | Mean :1.738 |
| ## | 3rd Qu.:14.610 | 3rd Qu.:0.11000 | 3rd Qu.:13.857 | 3rd Qu.:2.353 |
| ## | Max. :28.520 | Max. :0.55000 | Max. :30.660 | Max. :5.230 |
| ## | AARSD1 | AASDH | AASDHPPT | AASS |
| ## | Min. : 1.480 | Min. : 0.640 | Min. : 1.610 | Min. :0.0100 |
| ## | 1st Qu.: 5.388 | 1st Qu.: 2.667 | 1st Qu.: 6.850 | 1st Qu.:0.1100 |
| ## | Median : 8.795 | Median : 3.810 | Median : 9.545 | Median :0.2000 |
| ## | Mean : 9.754 | Mean : 4.128 | Mean :10.372 | Mean :0.2624 |
| ## | 3rd Qu.:13.178 | 3rd Qu.: 5.228 | 3rd Qu.:13.057 | 3rd Qu.:0.3475 |
| ## | Max. :22.520 | Max. :11.870 | Max. :27.650 | Max. :1.1100 |
| ## | AATF | AATK | ABAT | ABCA1 |
| ## | Min. : 7.20 | Min. : 1.230 | Min. : 0.410 | Min. : 0.320 |
| ## | 1st Qu.:31.05 | 1st Qu.: 4.185 | 1st Qu.: 3.825 | 1st Qu.: 4.157 |
| ## | Median :35.83 | Median : 5.865 | Median : 6.025 | Median : 9.750 |
| ## | Mean :36.95 | Mean : 6.451 | Mean : 7.035 | Mean :12.097 |
| ## | 3rd Qu.:43.55 | 3rd Qu.: 8.065 | 3rd Qu.: 9.885 | 3rd Qu.:17.767 |
| ## | Max. :56.99 | Max. :20.710 | Max. :18.910 | Max. :48.140 |
| ## | ABCA10 | ABCA12 | ABCA13 | ABCA2 |
| ## | Min. :0.0100 | Min. :0.0000000 | Min. :0.0100 | Min. : 0.470 |
| ## | 1st Qu.:0.1200 | 1st Qu.:0.0000000 | 1st Qu.:0.0825 | 1st Qu.: 5.080 |
| ## | Median :0.2100 | Median :0.0000000 | Median :0.2600 | Median : 6.945 |
| ## | Mean :0.2322 | Mean :0.0001587 | Mean :0.7864 | Mean : 7.667 |
| ## | 3rd Qu.:0.2900 | 3rd Qu.:0.0000000 | 3rd Qu.:0.9475 | 3rd Qu.: 9.457 |

| | | | | |
|----|----------------|-------------------|-----------------|-----------------|
| ## | Max. :0.6600 | Max. :0.0100000 | Max. :8.1000 | Max. :25.510 |
| ## | ABCA3 | ABCA4 | ABCA5 | ABCA6 |
| ## | Min. :0.040 | Min. :0.00000 | Min. :0.310 | Min. :0.000 |
| ## | 1st Qu.:0.210 | 1st Qu.:0.00000 | 1st Qu.:1.160 | 1st Qu.:0.030 |
| ## | Median :0.370 | Median :0.00000 | Median :1.950 | Median :0.065 |
| ## | Mean :0.421 | Mean :0.01667 | Mean :2.047 | Mean :0.109 |
| ## | 3rd Qu.:0.580 | 3rd Qu.:0.00000 | 3rd Qu.:2.690 | 3rd Qu.:0.120 |
| ## | Max. :1.250 | Max. :1.51000 | Max. :6.270 | Max. :2.670 |
| ## | ABCA7 | ABCA8 | ABCA9 | ABCB1 |
| ## | Min. : 1.50 | Min. :0.0000000 | Min. :0.0300 | Min. :0.0500 |
| ## | 1st Qu.:17.52 | 1st Qu.:0.0000000 | 1st Qu.:0.1100 | 1st Qu.:0.3675 |
| ## | Median :23.02 | Median :0.0000000 | Median :0.1550 | Median :0.9950 |
| ## | Mean :24.35 | Mean :0.0003968 | Mean :0.1785 | Mean :1.4771 |
| ## | 3rd Qu.:30.82 | 3rd Qu.:0.0000000 | 3rd Qu.:0.2075 | 3rd Qu.:2.1050 |
| ## | Max. :54.85 | Max. :0.0100000 | Max. :0.6800 | Max. :8.7800 |
| ## | ABCB10 | ABCB11 | ABCB4 | ABCB5 |
| ## | Min. : 1.490 | Min. :0.0300 | Min. :0.0000 | Min. :0.00000 |
| ## | 1st Qu.: 4.862 | 1st Qu.:0.0900 | 1st Qu.:0.1000 | 1st Qu.:0.04000 |
| ## | Median : 7.175 | Median :0.1300 | Median :0.1900 | Median :0.06000 |
| ## | Mean : 7.485 | Mean :0.1469 | Mean :0.2753 | Mean :0.07111 |
| ## | 3rd Qu.: 9.408 | 3rd Qu.:0.1800 | 3rd Qu.:0.3600 | 3rd Qu.:0.09000 |
| ## | Max. :19.300 | Max. :0.4300 | Max. :1.6500 | Max. :0.22000 |
| ## | ABCB6 | ABCB7 | ABCB8 | ABCB9 |
| ## | Min. :0.750 | Min. : 1.070 | Min. :0.2400 | Min. :0.0200 |
| ## | 1st Qu.:1.770 | 1st Qu.: 2.953 | 1st Qu.:0.8625 | 1st Qu.:0.1725 |
| ## | Median :2.485 | Median : 3.925 | Median :1.4600 | Median :0.3300 |
| ## | Mean :2.687 | Mean : 4.301 | Mean :1.7025 | Mean :0.5910 |
| ## | 3rd Qu.:3.265 | 3rd Qu.: 5.513 | 3rd Qu.:2.3000 | 3rd Qu.:0.6375 |
| ## | Max. :7.040 | Max. :10.350 | Max. :5.4300 | Max. :8.3300 |
| ## | ABCC1 | ABCC10 | ABCC11 | ABCC12 |
| ## | Min. : 1.070 | Min. :0.770 | Min. :0.00000 | Min. :0.00000 |
| ## | 1st Qu.: 3.745 | 1st Qu.:2.373 | 1st Qu.:0.01000 | 1st Qu.:0.00000 |
| ## | Median : 5.445 | Median :3.220 | Median :0.03000 | Median :0.00000 |
| ## | Mean : 5.967 | Mean :3.591 | Mean :0.08524 | Mean :0.00119 |
| ## | 3rd Qu.: 7.600 | 3rd Qu.:4.460 | 3rd Qu.:0.09000 | 3rd Qu.:0.00000 |
| ## | Max. :14.040 | Max. :9.640 | Max. :0.77000 | Max. :0.02000 |
| ## | ABCC2 | ABCC3 | ABCC4 | ABCC5 |
| ## | Min. :0.110 | Min. : 0.670 | Min. : 0.290 | Min. : 0.870 |
| ## | 1st Qu.:0.965 | 1st Qu.: 3.100 | 1st Qu.: 1.700 | 1st Qu.: 7.367 |
| ## | Median :1.485 | Median : 4.695 | Median : 2.990 | Median :10.950 |
| ## | Mean :1.638 | Mean : 5.295 | Mean : 3.902 | Mean :12.265 |
| ## | 3rd Qu.:2.257 | 3rd Qu.: 6.718 | 3rd Qu.: 5.088 | 3rd Qu.:16.782 |
| ## | Max. :5.110 | Max. :18.910 | Max. :13.850 | Max. :32.180 |
| ## | ABCC6 | ABCC8 | ABCC9 | ABCD1 |
| ## | Min. :0.110 | Min. :0.0000000 | Min. :0.130 | Min. : 0.800 |
| ## | 1st Qu.:2.120 | 1st Qu.:0.0000000 | 1st Qu.:1.195 | 1st Qu.: 2.450 |
| ## | Median :3.185 | Median :0.0000000 | Median :1.545 | Median : 3.125 |
| ## | Mean :3.434 | Mean :0.0006349 | Mean :1.692 | Mean : 3.482 |
| ## | 3rd Qu.:4.718 | 3rd Qu.:0.0000000 | 3rd Qu.:2.040 | 3rd Qu.: 4.357 |
| ## | Max. :8.960 | Max. :0.0300000 | Max. :5.080 | Max. :11.260 |
| ## | ABCD2 | ABCD3 | ABCD4 | ABCE1 |
| ## | Min. :0.0100 | Min. : 1.400 | Min. : 2.470 | Min. : 0.780 |
| ## | 1st Qu.:0.3025 | 1st Qu.: 4.645 | 1st Qu.: 5.768 | 1st Qu.: 4.463 |
| ## | Median :0.6800 | Median : 6.915 | Median : 8.365 | Median : 8.215 |

| | | | | |
|----|-----------------|-----------------|------------------|-----------------|
| ## | Mean :0.9944 | Mean : 7.178 | Mean : 8.666 | Mean : 9.527 |
| ## | 3rd Qu.:1.3350 | 3rd Qu.: 9.082 | 3rd Qu.:11.088 | 3rd Qu.:12.375 |
| ## | Max. :6.1300 | Max. :16.380 | Max. :20.000 | Max. :37.870 |
| ## | ABCF1 | ABCF2 | ABCF2-H2BE1 | ABCF3 |
| ## | Min. : 2.690 | Min. : 0.870 | Min. : 0.310 | Min. : 3.36 |
| ## | 1st Qu.: 8.777 | 1st Qu.: 3.045 | 1st Qu.: 2.945 | 1st Qu.:10.78 |
| ## | Median :12.980 | Median : 4.315 | Median : 4.820 | Median :14.12 |
| ## | Mean :14.384 | Mean : 4.759 | Mean : 5.534 | Mean :15.12 |
| ## | 3rd Qu.:18.567 | 3rd Qu.: 6.537 | 3rd Qu.: 7.080 | 3rd Qu.:18.55 |
| ## | Max. :33.600 | Max. :11.280 | Max. :18.930 | Max. :31.68 |
| ## | ABCG1 | ABCG2 | ABCG4 | ABCG5 |
| ## | Min. : 1.060 | Min. :0.00000 | Min. :0.000000 | Min. :0.00000 |
| ## | 1st Qu.: 7.643 | 1st Qu.:0.03000 | 1st Qu.:0.000000 | 1st Qu.:0.00000 |
| ## | Median :12.510 | Median :0.07000 | Median :0.000000 | Median :0.00000 |
| ## | Mean :15.407 | Mean :0.09825 | Mean :0.005079 | Mean :0.00254 |
| ## | 3rd Qu.:22.023 | 3rd Qu.:0.14000 | 3rd Qu.:0.000000 | 3rd Qu.:0.00000 |
| ## | Max. :44.820 | Max. :0.74000 | Max. :0.070000 | Max. :0.10000 |
| ## | ABCG8 | ABHD1 | ABHD10 | ABHD11 |
| ## | Min. :0.00000 | Min. :0.0000 | Min. : 2.700 | Min. : 1.510 |
| ## | 1st Qu.:0.00000 | 1st Qu.:0.0500 | 1st Qu.: 6.893 | 1st Qu.: 4.925 |
| ## | Median :0.01000 | Median :0.1100 | Median : 9.605 | Median : 6.770 |
| ## | Mean :0.01627 | Mean :0.1386 | Mean :10.167 | Mean : 6.907 |
| ## | 3rd Qu.:0.02750 | 3rd Qu.:0.1875 | 3rd Qu.:12.770 | 3rd Qu.: 8.258 |
| ## | Max. :0.06000 | Max. :0.4700 | Max. :21.830 | Max. :16.820 |
| ## | ABHD12 | ABHD12B | ABHD13 | ABHD14A |
| ## | Min. : 1.760 | Min. :0.0000 | Min. : 2.250 | Min. : 0.440 |
| ## | 1st Qu.: 4.570 | 1st Qu.:0.2500 | 1st Qu.: 8.315 | 1st Qu.: 2.513 |
| ## | Median : 6.975 | Median :0.5400 | Median :12.580 | Median : 4.365 |
| ## | Mean : 7.582 | Mean :0.6308 | Mean :12.925 | Mean : 5.992 |
| ## | 3rd Qu.:10.018 | 3rd Qu.:0.9425 | 3rd Qu.:17.340 | 3rd Qu.: 8.738 |
| ## | Max. :17.600 | Max. :2.1700 | Max. :28.730 | Max. :24.990 |
| ## | ABHD14A-ACY1 | ABHD14B | ABHD15 | ABHD16A |
| ## | Min. :0.00000 | Min. : 3.16 | Min. : 0.480 | Min. :16.15 |
| ## | 1st Qu.:0.00000 | 1st Qu.: 8.96 | 1st Qu.: 2.112 | 1st Qu.:39.33 |
| ## | Median :0.00000 | Median :15.04 | Median : 3.255 | Median :49.42 |
| ## | Mean :0.02397 | Mean :18.32 | Mean : 3.870 | Mean :48.98 |
| ## | 3rd Qu.:0.00000 | 3rd Qu.:24.36 | 3rd Qu.: 5.032 | 3rd Qu.:56.97 |
| ## | Max. :1.08000 | Max. :62.03 | Max. :10.410 | Max. :78.40 |
| ## | ABHD16B | ABHD17A | ABHD17B | ABHD17C |
| ## | Min. :0.000 | Min. : 1.710 | Min. : 0.640 | Min. :0.310 |
| ## | 1st Qu.:0.050 | 1st Qu.: 5.093 | 1st Qu.: 4.832 | 1st Qu.:1.195 |
| ## | Median :0.100 | Median : 7.610 | Median : 6.170 | Median :1.875 |
| ## | Mean :0.139 | Mean : 9.843 | Mean : 6.578 | Mean :2.102 |
| ## | 3rd Qu.:0.190 | 3rd Qu.:12.273 | 3rd Qu.: 8.777 | 3rd Qu.:2.730 |
| ## | Max. :1.190 | Max. :45.630 | Max. :13.290 | Max. :9.030 |
| ## | ABHD18 | ABHD2 | ABHD3 | ABHD4 |
| ## | Min. : 0.670 | Min. : 4.25 | Min. : 6.29 | Min. : 5.45 |
| ## | 1st Qu.: 2.705 | 1st Qu.:24.13 | 1st Qu.: 54.79 | 1st Qu.:24.17 |
| ## | Median : 4.050 | Median :39.67 | Median : 91.78 | Median :31.27 |
| ## | Mean : 4.253 | Mean :40.06 | Mean : 95.24 | Mean :30.76 |
| ## | 3rd Qu.: 5.720 | 3rd Qu.:54.10 | 3rd Qu.:129.38 | 3rd Qu.:37.72 |
| ## | Max. :10.040 | Max. :92.47 | Max. :209.00 | Max. :53.29 |
| ## | ABHD5 | ABHD6 | ABHD8 | ABI1 |
| ## | Min. : 7.75 | Min. :0.230 | Min. :0.330 | Min. : 6.07 |

```
## 1st Qu.: 37.02    1st Qu.:1.885    1st Qu.:1.048    1st Qu.:35.78
## Median : 56.00    Median :2.930    Median :1.565    Median :53.73
## Mean   : 64.62    Mean   :3.104    Mean   :1.818    Mean   :51.73
## 3rd Qu.: 82.17    3rd Qu.:4.117    3rd Qu.:2.317    3rd Qu.:67.18
## Max.   :181.54    Max.    :8.460    Max.    :5.490    Max.    :99.61
##      ABI2
## Min.    :0.140
## 1st Qu.:1.075
## Median :2.005
## Mean    :2.323
## 3rd Qu.:3.203
## Max.    :8.680
```

```
library(ggplot2)
```

```
# Making my histogram for A1BG expression
```

```
ggplot(merged_proj_data, aes(x = A1BG)) +
```

```
# Decided to make my binwidth 0.05 for more data points/make the data points more distinct
```

```
# Also customized my colors
```

```
geom_histogram(binwidth = 0.05, fill = "#D1514D", color = "black") +
```

```
# Added a title, and x and y axis labels
```

```
labs(title = "A1BG Expression",
```

```
      x = "A1BG Expression", y = "Count") +
```

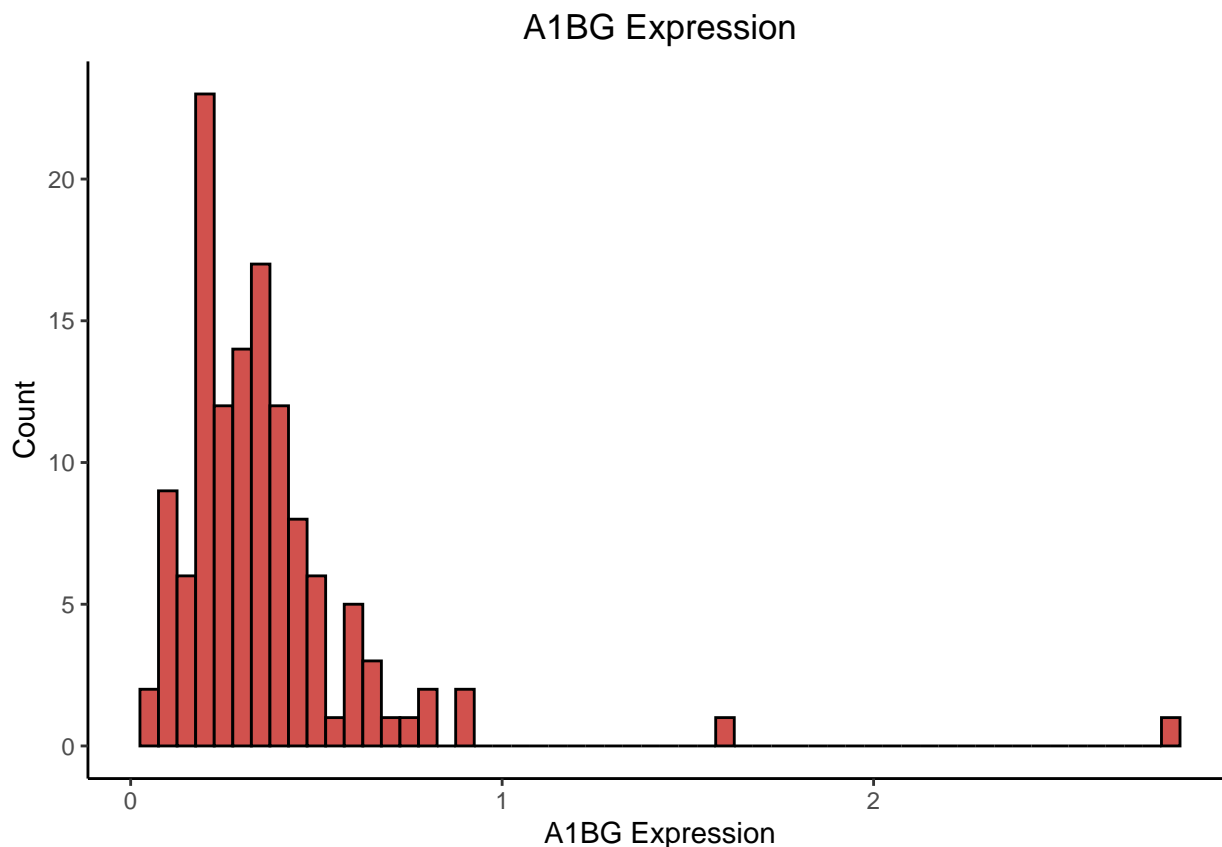
```
# Set the theme to classic for a clean background
```

```
theme_classic() +
```

```
# Wanted to center my title
```

```
##https://www.r-bloggers.com/2025/03/how-to-center-ggplot-title-subtitle-and-caption-in-ggplot2-wi
```

```
theme(plot.title = element_text(hjust = 0.5))
```

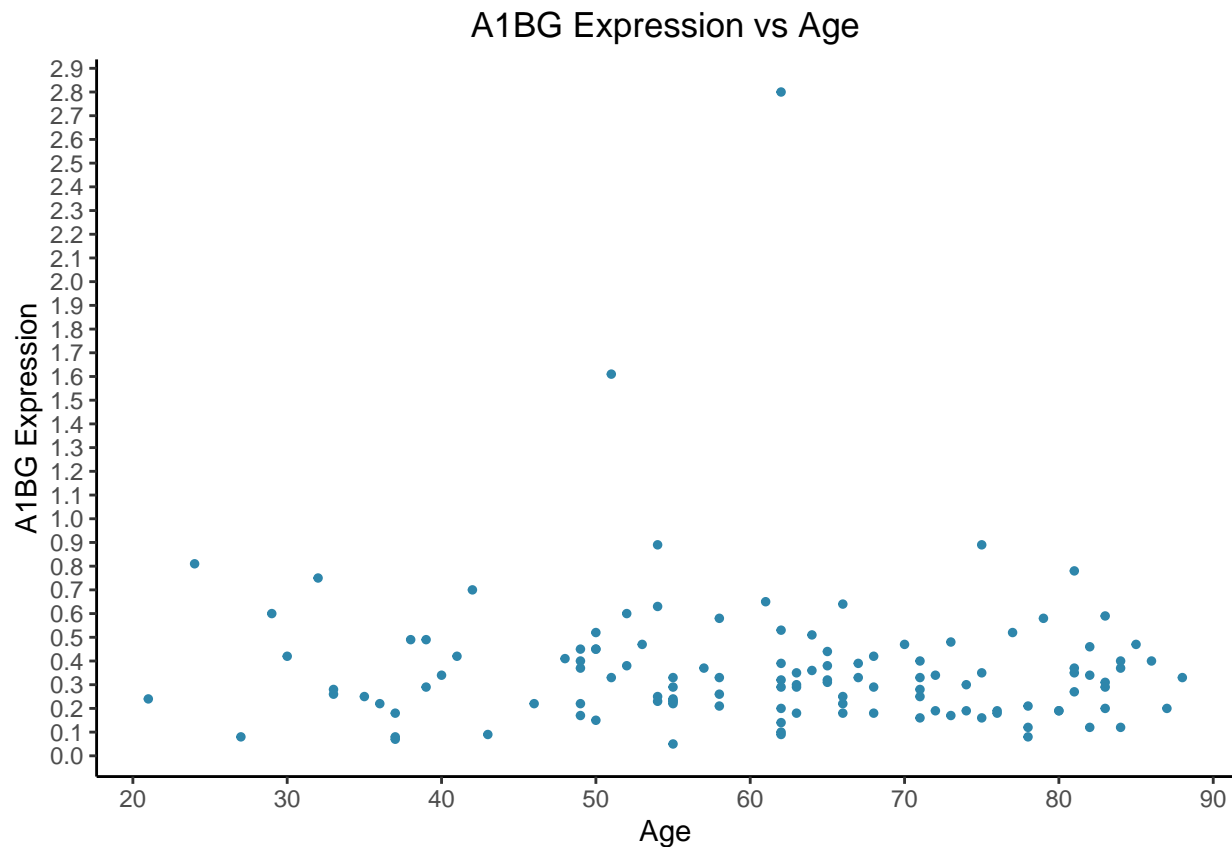


```
# "age" was not being registered as a number, so I had to change the age column to numeric
merged_proj_data$age <- as.numeric(merged_proj_data$age)
```

```
## Warning: NAs introduced by coercion
```

```
# Making my scatterplot for A1BG Expression vs Age
ggplot(merged_proj_data, aes(x = age, y = A1BG)) +
  # Customized the color and the size -- I set the size to 1 so that the point is more precise (when it
  geom_point(color = "#2E86AB", size = 1) +
  # Added a title, and x and y axis labels
  labs(title = "A1BG Expression vs Age",
        x = "Age", y = "A1BG Expression") +
  # Customized the scale for x axis to go up by 10s to make the plot easier to read
  ##https://www.sthda.com/english/wiki/ggplot2-axis-scales-and-transformations#google_vignette
  scale_x_continuous(breaks = seq(0, 100, by = 10)) +
  # Customized the scale for y axis to go up by 0.1 to make the plot easier to read
  scale_y_continuous(breaks = seq(0, 3, by = 0.1)) +
  # Set my theme to classic for a clean background
  theme_classic() +
  # Centered the title
  theme(plot.title = element_text(hjust = 0.5))
```

```
## Warning: Removed 3 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

```
# Making my boxplots for A1BG expression by disease status and sex
ggplot(merged_proj_data, aes(x = disease_status, y = A1BG, fill = sex)) +
  geom_boxplot() +
  # Customized the colors for sex
  scale_fill_manual(values = c("male" = "#849FE1", "female" = "#D1514D", "unknown" = "#E0C06B")) +
  # Added a title, and x and y labels
  labs(title = "A1BG Expression by Disease Status and Sex",
        x = "Disease Status", y = "A1BG Expression") +
  # Set my theme to classic for a clean background
  theme_classic() +
  # Centered my title
  theme(plot.title = element_text(hjust = 0.5))
```

